DOCUMENT RESUME

ED 399 000 PS 024 236

AUTHOR	Hau, Kit-Tai
TITLE	Stage Structure of Moral Development: A Comparison of Alternative Models.
PUB DATE	Apr 96
NOTE	24p.; Paper presented at the Annual Conference of the American Educational Research Association (New York, NY, April 8-12, 1996).
PUB TYPE	Speeches/Conference Papers (150) -- Reports - Research/Technical (143)
EDRS PRICE	MF01/PC01 Plus Postage.
DESCRIPTORS	*Adolescents; *Adults; Comparative Analysis; Developmental Stages; Foreign Countries; *Goodness of Fit; *Models; *Moral Development; Moral Values; Structural Equation Models; Value Judgment
IDENTIFIERS	British; Chinese People; Kohlberg (Lawrence); Moral Development Test; Moral Orientation

ABSTRACT

This study evaluated the stage structure of several quasi-simplex and non-simplex models of moral development in two domains of moral development in a British and a Chinese sample. Analyses were based on data reported by Sachs (1992): the Chinese sample consisted of 1,005 students from grade 9 to post-college, and the British sample consisted of 60 students in grades 9-12 and adults. Subjects completed the Moral Development Test, which assesses affective (moral orientation) and cognitive (moral judgment) aspects of moral development. The alternative simplex models evaluated in the present study involved minor reversals, moderate reversals or serious reversals of stages. Ten non-simplex models considering alternate paths of development were also evaluated. The models were evaluated using goodness of fit with the Tucker Lewis Index and the Relative Noncentrality Index and the extent to which the model converged to solutions with reasonable parameter estimates. A strong alternative model would be one that was applicable to both moral orientation and moral judgment and to both cultural groups. The fit indexes and the parameter estimates revealed no strong competitor for the theoretical model. (Contains 34 references.) (KDFB)

Stage Structure of Moral Development:

A Comparison of Alternative Models

Kit-Tai Hau

The Chinese University of Hong Kong

Paper presented at the American Educational Research Association Annual Conference, New York, April 1996. This research was supported in part by the Hong Kong Institute of Educational Research and the Shaw College both at the Chinese Univeristy of Hong Kong. Correspondence concerning this article should be addressed to Kit-Tai Hau, Faculty of Educaiton, The Chinese Univeristy of Hong Kogn, Shatin, N.T., Hong Kong (e-mail: KTHAU@CUHK.EDU.HK)

Abstract

Previous structural equation analyses on moral stage sequence usually evaluated only the fit of the theoretical model (e.g., Stages 2->3->4->5) to the data. This strategy is relatively weak because of the existence of other equivalent or plausible alternative models which might have identical, or even better fit. The present study evaluated an exhaustive list of simplex models as well as a number of other plausible alternative models in two domains of moral development over two cultural samples (Chinese and British). Results revealed no strong competitors to the theoretical model. Issues related to simplex structure and the importance of examining alternative models were also discussed.

Stage Structure of Moral Development: A Comparison of Alternative Models

The stage sequence and structure in moral development is of great interest to social psychologists (e.g., Kohlberg, 1969; Rest, 1986). Using structural equation modelling, previous researchers (e.g., Boom & Molenaar, 1989; Sachs, 1992) fit their data to the theoretical model and found some support for the linear development from lower to higher stages (e.g., from Stage 2 to 3, 4, 5, and 6). However, it is possible that other stage sequences or alternative models will fit and explain the data equally well, or even better. The present study re-examined the issue using an exhaustive list of simplex models as well as a number of other closely related non-simplex models. These alternative models were evaluated by their fit to the data and properness of the parameter estimates.

Moral Development

In the study of moral development, the late Lawrence Kohlberg has been considered the only contemporary psychologist who embraces philosophy as important and essential in the definition of what is moral. He posited a 3 levels and 6 stages hierarchical model (2 stages within each level) which composed of the stages: heteronomous (Stage 1; e.g., avoidance of punishment), individualism (e.g., reward), mutual interpersonal expectations, social system (law and order), social contract, and universal principles (Stage 6) (Rest, 1983).

Results from cross-cultural studies (for reviews, see Rest, 1986; Snarey, 1985) generally supported the invariant structure of lower moral stages. However, studies using Kohlberg's Moral Judgment Interview (MJI) showed that the moral development in other cultures, especially peasant villages, was usually slower than that in Kohlberg's U.S. norms (Snarey, 1985). Noteworthily, in some studies (Gorsuch & Barnes, 1973; White, Bushnell & Regnemer, 1978) moral reasoning beyond stage 3 was totally absent at the age of 16. The invariance of higher stages was not unanimously supported (Edwards, 1975; Maqsud, 1979).

Davison and his colleagues (1977, Davison, Robbins & Swanson, 1978) proposed the use of a metric unfolding model to test the hierarchical stage structure. It was posited that a principal components analysis of the stage scores should produce two factors. The first factor should have the highest

loadings in the intermediate stages while the second factor should have loadings reproducing the order of the stages. Davison et al. showed that their subjects displayed a moral development sequence matching the theoretical one.

Except the original work by Davison et al., however, little similar research has been published on the reproduction of stage structure with other samples, especially in a cross-cultural setting. Using a 4-story short version of the Defining Issues Test (DIT, Rest, 1979), a most popular objective moral judgement test (Kay, 1982), Ma (1985, 1988) found that the factor structures as produced by the British subjects were unstable across samples and were sometimes inconsistent with the theoretical stage order. Furthermore, the order of stages 4, 5, and 6 of Chinese in Hong Kong and Mainland China were quite non-discriminating (Ma, 1988; Ma & Chan, 1987). The findings also suggested a cultural difference in that the Chinese tended to perceive stage 4 reasoning as more similar to stages 5 and 6 rather than to stages 2 and 3 (Ma 1988, Ma & Chan, 1987). Hau and Lew (1989) found similarly that the sequence between levels (e.g., between levels I and II) were quite distinct, whereas those within a level (e.g., between stages 5 and 6 in Level III) were rather ambiguous.

<u>**Simplex Structure**</u>

Simplex models have been used widely to examine longitudinal data in which the same variable is measured repeatedly on the same people over several occasions (Joreskog & Sorbom, 1988; Marsh, 1993). Davison (1977, Davison et al., 1978) argued that the correlations of the stage scores should also display a simplex-like pattern, that is, in any row or column, the correlations should fall off when one moves away from the main diagonal. This is a reflection of the fact that the sizes of correlations between adjacent stage scores (e.g., Stages 3 and 4) are larger than those further apart (e.g., Stages 3 and 5). Actually the correlation between any two non-adjacent stages (e.g., Stages 3 and 5) is zero when the effect due to an in between stage (e.g., Stage 4) is partialled out (Joreskog, 1970; Marsh, 1993).

Consider the path model involving 4 stages (Stages 2, 3, 4, and 5) (see Figure 1). The rectangular boxes (y_i) represent the observed stage scores (usually means of items for a particular stage), whereas ovals (h_i) are the latent constructs reflecting the true stage scores, the e s are the

errors in measurement, the z s are factor residuals, the l s are the factor loadings between observed stage scores and the corresponding true (latent) stage scores, and b s are the path coefficient between latent stage scores which reflect the proximities of consecutive stages.

Insert Figure 1 about here

The perfect simplex model is distinguished from the quasi-simplex one in that the stage scores of the former are assumed to be measured without error (Joreskog, 1970; Joreskog & Sorbom, 1988). In that case, all e s are zero with the l s being fixed to one. In contrast, the quasi-simplex model to be used in the present study hypothesizes that the observed stage scores contain a measurement error component.

$y_t = h_t + e_t$ for $t = 1$ to T occasions

$h_t = b_t h_{t-1} + z_t$ for $t = 2$ to T occasions.

Despite that the quasi-simplex model allows measurement errors and reflects a closer picture of real empirical data, there are identification problems in such model. It can be shown mathematically that the parameter estimates in the first and last two stage scores cannot be uniquely identified. For the model in Figure 1, one indeterminacy is associated with b_2, y_1, y_2 and q_1 ; while the other involves y_4 and q_4 (Joreskog, 1970; Joreskog & Sorbom, 1988, pp.182-186)[where $y = \mathrm{Var}(z)$ and $q = \mathrm{Var}(e)$]. One condition must be imposed on each of these two sets of parameters to eliminate the indeterminacies.

In earlier works (e.g., Joreskog, 1970, 1977), the above indeterminacy was solved by fixing the error terms (q_1 and q_4) at both ends to be zero. This method was adopted by Boom and Molenaar (1989) and Sachs (1992) in their analyses. However, in view of the fact that moral judgement measures, similar to a lot of other psycho-social indicators, have only moderate reliabilities (.27 to .78 for individual stage scores; Davison & Robbins, 1978; Hau & Lew, 1989; Rest, 1979), it would be unreasonable to set q_1 and q_4 to be zero, which in effect implies no measurement errors in the stage scores at the two ends. In more recent works, Joreskog (1981, Joreskog & Sorbom, 1988) suggested,

"The most natural way of eliminating the indeterminacies is to set $q_1 = q_2$ and $q_3 = q_4$" (1988, p.186). One purpose of the present study is to reanalyse some of the results of previous studies using the latter strategy which allows measurement errors in all stage scores. Noteworthily, it can be shown empirically that the above two methods of solving indeterminacy lead to the same goodness of fit to the data, whereas the parameter estimates can be substantial different (see Table 1).

In the analyses of longitudinal data using simplex models, Marsh and Hau (1994) have also noted that it is necessary to include a priori correlated uniquenesses relating the same indicators administered on different occasions. In the models being examined in present study, correlated uniqueness terms were not included. This is justified because unlike longitudinal study, there are no common indicators across different stages. It is believed that the effects due to correlated uniqueness, if any, would be relatively small.

Alternative and Equivalent Models

In the evaluation of structural equation models, Bollen (1989, pp.67-72) and many others (e.g., MacCallum, Wegener, Uchino, and Fabrigar, 1993) criticize that there is frequent confusion between model-data and model-reality consistency. The former checks whether the data is consistent with the data, whereas the latter examines whether the model is consistent with the real world. The link between the two is asymmetric in that if the data are consistent with a model, it does not necessarily follows that the model reflects the reality. The existence of other competitive models which have identical or even better fit to the data cannot be ignored. Rather, they have to be eliminated either empirically (showing they do not fit the data) or theoretically (showing they are not logically possible).

In the following we will first discuss equivalent models which have the same goodness of fit. Then we will examine other alternative models which may have worse or better fit to the data. As regards equivalent models, a trivial case is that the theoretical model and one with exactly opposite sequence (e.g., Stages 5 -> 4 -> 3 -> 2 instead of Stages 2 -> 3 -> 4 -> 5) have identical fit to the data. As we are fixing the q s of the first two stages to be equal and those of the last two stages to be equal, it can be shown empirical that in this particular model, a reversal of order of the respective two stages

will not affect the goodness of fit (e.g., Stages 2 -> 3 -> 4 -> 5, Stages 3 -> 2 -> 4 -> 5, Stages 3 -> 2 -> 5 -> 4 have the same fit to the data, see Table 1). These alternative models can only be distinguished and eliminated by examining whether they all converge to proper solutions, which include positive measurement errors (q s) and factor residuals (y s), reasonable standard errors, and standardized regression (b s) smaller than one. Furthermore, in the simplex model, as b s reflect the proximities between consecutive stages, they are hypothesized to take positive values only.

Other possible alternative models include those that are slightly (e.g., involving reversal of two adjacent stages) or grossly different (e.g., involving reversal of non-adjacent stages) from the theoretical sequence. In the present study with 4 stages (Stages 2, 3, 4, and 5), there are 24 possible combinations of stage order in the simplex model, some of which are equivalent to each other in their fit to the data.

A lot of other alternative non-simplex models can also be generated which might have better or worse fit to the data. An example is to have Stages 4 and 5 as alternative paths for development. That is, all children progress from stages 2 to 3, but some will proceed to stage 4 and end there while others go on to stage 5 directly without passing through stage 4 [represented by: stages 2 -> 3 -> (4, 5)] (see M2 in Table). This model in efeect is arguing that stages 4 and 5 are alternate paths of moral maturation.

Sachs (1992) investigated and compared the hierarchical stage order of the Moral Development Test (MDT) (Ma, 1987) over two cultures (British and Chinese). However, he inspected only the fit of the data to the theoretical model (stages 2 -> 3 -> 4 -> 5). The main purpose of the present study was to re-examine the stage sequence by inspecting an exhaustive list of simplex models as well as a number of other plausible non-simplex alternatives. These models were evaluated by their fit to the data together with the properness of various parameter estimates in the model.

Method

The analyses in the present study were based on data collected by Ma (1988, 1989) and reported by Sachs (1992). Subjects from two cultures were examined. The 1005 subjects in the Chinese sample consisted of 90 Grade (G.) 9, 188 G.10, 243 G.11, 164 G.12, 302 college students and

18 adults from Hong Kong, whereas the 281 British subjects consisted of 60 G.9, 112 G.10, 61 G.12 students and 48 adults. There were approximately equal number of male and female subjects in each sample.

All subjects completed the Moral development Test (MDT) which assessed both the affective (moral orientation) and cognitive (moral judgment) aspects of moral development (Ma, 1988, 1989). The moral orientation (N scores) scales were used to measure subjects' tendency to gratify psychological needs and to perform altruistic acts towards others (Sachs, 1992). Whereas, the moral judgment (J scores) scales were parallel to Rest's Defining Issues Test (DIT) and measured subjects' cognitive maturity in moral judgment as defined by Kohlberg (1969).

The four correlation matrices of the stage scores reported by Sachs (1992) on the two domains (moral orientation and moral judgment) and for the two samples (Chinese and British) served as input matrices for the following analyses. Despite the fact that correlation rather than covariance matrices were used, this did not affect substantially our results.

First, in the following particular analyses, goodness of fit indexes were identical irrespective of whether correlation or covariance was used. Second, because all indicators were actually means of sets of items on the same Likert scale (e.g., 7 points), it is reasonable to assume that their means and variances would not differ too much. Actually re-analyses with indicators having slightly different means and variances showed that the following conclusions were robust to such variation. Third, the setting of equal measurement errors of the first two and last two stage scores in the correlation matrices was effectively assuming that the respectively scales had similar reliability (or proportion of error variance to true variance). Whereas, if covariance matrices had been used, the setting of equal q s was equivalent to the assumption that the absolute error variances (i.e., in the original metric) of the scales were identical. Apparently, in this particular analysis (especially when we are not setting further constraints across cultural groups), the former assumption of using correlation matrices was as good as, if not better than, that basing on covariance matrices.

Results

Model Evaluation

When evaluating the goodness of fit of the models, the TLI (Tucker Lewis Index) and RNI (Relative Noncentrality Index) were chosen because unlike many others, these two indices are unbiased in that their expected values do not vary systematically with sample size. The former, however, differs in that it embodies a control for model complexity and a reward for parsimony. Both incremental indexes have been recommended for routine use in model evaluation (Marsh, Balla, & Hau, 1994).

Models were also checked to see that they converged to solutions with reasonable parameter estimates, which included: nonnegative measurement errors (q s), nonnegative factor residuals (y s), standardized regression paths (b) which were positive but less than one. The results of the evaluation for the two moral domains (orientation and judgment) and in the two cultural samples are shown in Table 1. When a certain model is improper, one of the sources of problems together with the respective illegal values were also tabulated (e.g., q = TE = -.87). Sometimes, the parameter estimates may fall on the boundary of legitimate domain. These solutions, though proper, may not be reasonable and interpretable (e.g., stage score with no measurement error, q = 0). They are labelled as worrisome solutions in the table.

It should be noted, however, that for some less serious problems (such as a negative uniqueness close to the zero boundary), it is possible that the analyses of the covariance matrices with stage scores having large differences in means and variances might bring the improper solutions back to the permissible range. However, a glance through the degree of improperness in Table 1 suggest that the majority of these improper solutions are unlikely to become proper irrespective of the differences in the means and variances of the stage scores.

<u>Simplex Models</u>

Preliminary analyses started with the comparison of simplex models formed by the two strategies ($q_1 = q_4 = 0$ and $q_1 = q_2$, $q_3 = q_4$) used to solve the parameter indeterminacy. As can be seen from Table 1, the two methods (M0 and M1) applied to the theoretical model (Stages 2->3->4->5) lead to the same goodness of fit but slightly different parameter estimates.

It was noticed that the solutions for the Chinese sample in the two moral domains were problematic. For the orientation scales, the two specifications used to solve parameter indeterminacy both resulted in improper solutions (due to negative measurement errors/uniquenesses of -.87). On the other hand, fitting the theoretical model to the Chinese moral judgment data gives rise to worrisome solutions of either zero measurement error or standardized regression path coefficient of one.

Twenty-three additional simplex models (M2 to M24 in Table 1) were generated for each domain and cultural sample. These models can be broadly classified as either minor, moderate, or serious reversal of the theoretical one. The minor reversals involved only one reversal of two adjacent stages, whereas in the moderate ones, a stage was misplaced two stages from its theoretical order (e.g., stage 2 came after stage 4). All other types of reversals, including a complete reversal of the theoretical model (i.e., stage 5->4->3->2) were grouped under the serious category. Noteworthily, all these models had the same degree of freedom (df=1).

A browse through the minor reversal category showed that the fits of these models were not particularly good. Actually all these 12 solutions either nonconverged or were improper. Similarly for the moderately and seriously reversed ones, only a few of the solutions were proper. Coincidentally, they were all in the British group.

For the three proper solutions (M11, M19, M24) of the British subjects in the moral orientation domain, the first two were equivalent in that one was just the complete reversal of the other and thus they had identical fit to the data. Despite the fact that they had comparable RNI (.98) with the theoretical model (M0, RNI=1.00), their chi-square and TLI (3.46 and .86 respectively) were much worse (for M0, Chi = .46, TLI = 1.03). Thus, there was no compelling reason to replace the theoretical model with these two alternative ones. M24 was just the complete reversal of the theoretical one and was expected mathematically to have identical fit to the data. Without other substantial empirical or theoretical support, there was no strong ground to accept this as a replacement for the theoretical one.

Three proper solutions (M5, M16, M24) were also found in the moral judgment domain of the British subjects. M5 and M16 were equivalent in that one was the complete reversal of the other. Both of them had goodness of fit as good as, if not better than, the original theoretical model (M5 and M16: Chi=1.42, TLI=.98, RNI=1.00 versus M0: Chi=1.81, TLI=.97, RNI=.99). M5 (stages 2->4->5->3) looked closer to the theoretical model whereas M16 (stages 3->5->4->2) involved greater reversals of stages. Both models can serve as potential competitors for the theoretical model. M24 was the complete reversal of the theoretical model and was also proper. As discussed above, there was no compelling reason to accept this as a replacement of the theoretical model.

A closer examination of the fit indexes in Table 1 shows that the 24 models are actually composed of three groups, each with eight models. Within each group, all models are equivalent and have identical goodness of fit. As constraints have been set on the error terms at the two ends ($q_1 = q_2$ and $q_3 = q_4$), it can be seen from Table 1 that models of the following stage sequence are equivalent: A->B->C->D; B->A->C->D; A->B->D->C; B->A->D->C; C->D->A->B; C->D->B->A; D->C->A->B; and D->C->B->A. Irrespective of the sample being used, all these models would have the same goodness of fit. However, it should be noted that the parameter estimates of these models can and will usually differ.

Non-Simplex Alternative Models

It is possible to generate a huge number of non-equivalent non-simplex alternative models as competitors for the theoretical model. For example, one can hypothesize that stages 3, 4, 5 are end-point themselves, but are alternative pathway from stage 2. That is, there are three regression paths going from stage 2 to stages 3, 4, 5 and there is no other path linking the stages. This is represented as 'stage 2->(2,3,4)' (see M1 in Table 2).

Due to the great number of possible alternative models, in the following analyses we are limiting ourselves to those that are closer to the theoretical model and have simple structure with only three paths linking the four stages (same number as the simplex model). Stage 2 is placed either as the first or among the earliest stages in the path diagram. Admittedly it is possible to have other grossly different models which have good fit to the data. However, as the theoretical model seems to

fit the data relatively well, it is reasonably to expect that the best fit alternative model would have a structure not much different from the theoretical one. Thus, the use of the present set of models, which closely resemble the theoretical one, as the starting point for exploration is well justified.

Using the criteria listed above, ten alternative non-simplex models were constructed (see Table 2). Similar to the indeterminacy problem in the simplex models, not all of the error terms (e1 to e4) could be separately estimated. The problem cannot be totally solved even by setting pairs of the error terms to be equal. Thus, in view of the assumption that the reliability of the stage scores would be approximately the same, it was decided to set all four error terms to be equal (i.e., $q_1 = q_2 = q_3 = q_4$). This solved the identification problem in all models except M4.

However, an inspection of the solutions for moral judgment in the two cultural groups showed that the factor residuals were on the boundary (-.06, -.07) (see M0A Table 2). This indicated either the theoretical model or the assumption of equal measurement errors might not be appropriate. This could not be verified without the support of other empirical data. Nonetheless, the model was reanalysed by setting e4 to be slightly smaller than the others ($q_1 = q_2 = q_3 = 1.05\, q_4$) (see M0B in Table 2). This led to proper solutions for both domains in the two cultural samples with no change in the goodness of fit. Noteworthily, reanalyses of other models (M1 to M10) using a slightly smaller e4 term did not substantially change the results.

An examination of the solutions in Table 2 shows that a number of them are proper. However, none of the models appeared to be strong competitor for the theoretical one. A lot of the proper solutions had very poor fit to the data (moral orientation, Chinese: M2, M9, M10, UK: M7, M8, M10; all proper models in moral judgment). Models M2 and M6 of the British subjects in the moral orientation domain had slightly larger chi-square, but comparable TLI and RNI to the theoretical model. Disappointedly, the same model when applied to other cultural samples or moral domain did not give the same kind of good fit. Nevertheless, M2 and M6 can still serve as potential alternative models in the future exploration of moral stage structure.

Discussion and Conclusion

The present study evaluated the stage structure of moral development in a wide range of quasi-simplex and non-simplex models. In the context of this study, a strong alternative model should be one that is applicable in both moral domains and is universal for both cultural groups. Despite the large number of models being inspected, an examination of the fit indexes as well as the parameter estimates revealed no strong competitor for the theoretical model. Even when we limit ourselves to good model for a particular moral domain and cultural sample, only one or two of the models could serve as potential competitors.

Despite the finding of no strong competitor, the conclusion that the theoretical model best describes moral development has to be taken with great caution. It is quite possible that a lot of the nonconvergent and improper solutions of competitive models were due to the reliability, or rather the lack of reliability, of the stage scores. This may lead to an unstable correlation matrix which do not fully conform to any simplex structure. That is, irrespective of how.we rearrange the stage order, the value in the correlation matrix do not systematically decrease as it moves away from the main diagonal. Or, the correlations between any two non-adjacent stages are not close to zero when the effect due to an in-between stage is partialled out.

The indeterminacy problem inherent with quasi-simple structures worsens the already slim chance of good fit because in effect, we have to force pairs of error terms to be equivalent. This, however, can be overcome by using multiple indicators for each stage. As advocated by Marsh (1993), this much stronger model allows each parameter to be estimated independently, or if necessary, restricted to be equal (or according to other constraints).

The use of multiple indicators may sometimes help to solve the problem of low reliability in stage score. For stages consisting of a number of coherent item parcels (composed of a set of unidimensional items), the aggregation of all items into one score may result in low reliability. However, when models with multiple indicators are used, the item parcels do not have to be aggregated. Each of these parcels can have their own contribution to the latent stage factor. This may lead to a more appropriate evaluation of different competitive stage structures.

As in other structural equation analyses, it is advisable to examine and verify the results of the present study using other established instruments and in additional samples and moral domains. Furthermore, due to moral immaturity, it is possible that the higher stages may appear to be non-differentiating to younger subjects. Thus, separate analyses for different age groups, as carried out by Boom and Molenaar (1989), will give additional support to the validity of any hypothesized stage structure.

In contrast to previous structural equation analyses on moral development (e.g., Boom & Molenaar, 1989; Sachs, 1992), the present study evaluates a great number of plausible alternative simplex and non-simplex models. This process of simultaneous examination and, perhaps, elimination of various alternative and plausible models is very important, but frequently neglected and even misinterpreted. For example, Randhawa, Beamer and Lunberg (1993) fit their cross-sectional data to one a priori model and postulated a causal structure among mathematics attitude, achievement and efficacy. In a comment on this research, Marsh et al. (1994) pointed out the existence of a number of equivalent models, including one with direction of causality just opposite to that hypothesized by Randhawa et al. Marsh et al. challenged Randhawa et al.'s conclusion because there was no basis to differentiate among these substantially different models. However, in responding to such criticism, Randhawa and Beamer (1994) clenched to their wrong belief and insisted, "Empirical equivalence is something that technicians can have fun with, but it is not the raison d'etre for theoretical repudiation" (p.465).

As exemplified by the above analyses, equivalent models cannot be distinguished in terms of their fit to the data. Rather, they can only be differentiated or eliminated by criteria such as interpretability of parameter estimates and meaningfulness of the model (MacCallum et al., 1993). Sometimes researchers defend their one model analysis and claim that their a priori model is more justifiable than other equivalent or alternative models because it is developed from prior research or theories. MacCallum et al. (1993) criticized and pointed out that "such defense would often be the product of wishful thinking" because "[the argument] implies that no other equally good explanation

of the data is as plausible as the researcher's a priori model simply because the researcher did not generate the alternatives a priori" (p.197).

In the above comparison of various competitive models, two remarks can also be made as regards model evaluation. First, by the standard of goodness of fit indexes alone, a lot of the models being evaluated are extremely good. However, more detail inspection of the parameter estimates reveal that most of these models are improper. In line with Bollen and Long's (1993, pp.6-7) recommendation, this points out the importance of examining the parameter estimates of various components in model evaluation.

Second, when the TLI and RNI indexes are compared, it can be seen that their respective values can differ significantly. For example, in the moral judgement simplex-model analyses for the Chinese sample (see Table 1), TLI is .93 when RNI is .99. But when TLI drops to .74 RNI is still .96. Again a further drop of TLI to .39 only results in a small drop of RNI to .90. It is clear that these two indexes may not be working on the same metric. Along with Bollen and Long (1993), Gerbing and Anderson, and Marsh, Balla and Hau (1994), we recommend that researchers should consider a number of appropriate indexes in model evaluation.

All in all, the present study evaluate a great number of competitive models of moral development. This process is important and can help to eliminate some possibilities or suggest alternative explanations. This is definitely not an number crunching game as some researchers may have misconceived (e.g., Randhawa & Beamer, 1994). Perhaps, Bollen's (1989, see also Bollen & Long, 1993, p.7) advice should be reiterated, "We need to examine other plausible specifications that fit; we need to explore various avenues to assess whether a model has a reasonable correspondence to reality" (p.72).

References

Bollen, K. A. (1989). Structural equations with latent variables. NY: John Wiley & Sons.

Bollen, K. A. & Long, J. S. (1993). Introduction. In K. A. Bollen & J. S. Long (Eds.), Testing structural equation models (pp.1-9). Newbury Park, CA: Sage.

Boom, J., & Molenaar, P. C. M. (1989). A developmental model of hierarchical stage structure in objective moral judgments. Developmental Review, 9, 133-145.

Davison, M. L. (1977). On a metric, unidimensional unfolding model for attitudinal and developmental data. Psychometrika, 42, 523-548.

Davison, M. L., & Robbins, S. (1978). The reliability and validity of objective indices of moral development. Applied Psychological Measurement, 2, 389-401.

Davison, M. L., Robbins, S., & Swanson, D. B. (1978). Stage structure in objective moral judgments. Developmental Psychology, 14, 137-146.

Edwards, C. P. (1975). Societal complexity and moral development: A Keyan study. Ethos, 3, 505-527.

Gerbing, D. W., & Anderson, J. C. (1993). Monte Carlo evaluations of goodness-of-fit indices for structural equation models. In K. A. Bollen & J. S. Long (Eds.), Testing structural equation models (pp.1-9). Newbury Park, CA: Sage.

Gorsuch, R. L., & Barnes, M. L. (1973). Stages of ethical reasoning and moral norms of Carib youths. Journal of Cross-cultural Psychology, 4, 282-301.

Hau, K. T. & Lew, W. J. F. (1989). Moral development of Chinese students in Hong Kong. International Journal of Psychology, 24, 561-569.

Joreskog, K. G. (1970). Estimation and testing of simplex models. British Journal of Mathematical and Statistical Psychology, 23, 121-145.

Joreskog, K. G. (1981). Analysis of covariance structure. Scandinavian Journal of Statistics, 8, 65-92.

Joreskog, K. G., & Sorbom, D. (1988). LISREL 7 - A guide to the program and applications (2nd ed.). Chicago, Illinois: SPSS.

Joreskog, K. G., & Sorbom, D. (1977). Statistical models and methods for analysis of longitudinal data. In D. J. Aigner and A. S. Goldberger (Eds.), Latent variables in socio-economic models (pp.285-325). Amsterdam: North-Holland.

Kay, S. R. (1982). Kohlberg's theory of moral development: Critical analysis of validation studies with the Defining Issues Test. International Journal of Psychology, 17, 27-42.

Kohlberg, L. (1969). Stage and sequence: The cognitive-developmental approach to socialization. In B. A. Goslin (Ed.), Handbook of socialization theory and research. San Franciso: Rand McNally.

Ma, H. K. (1985). Consistency of stage structure in objective moral judgment across different sample. Psychological Reports, 57, 987-990.

Ma, H. K. (1988). Objective moral judgment in Hong Kong, Mainland China, and England. Journal of Cross-Cultural Psychology, 19, 78-95.

Ma, H. K. (1989). Moral orientation and moral judgment in adolescents in Hong Kong, Mainland China and England. Journal of Cross-Cultural Psychology, 20(2), 152-177.

Ma, H. K., & Chan, W. S. (1987). The moral judgments of Chinese students. The Journal of Social Psychology, 127, 491-497.

MacCallum, R. C., Wegener, D. T., Uchino, B. N., & Fabrigar, L. R. (1993). The problem of equivalent models in applications of covariance structure analysis. Psychological Bulletin, 114, 185-199.

Maqsud, M. (1979). Cultural influences on transition in the development of moral reasoning in Nigerian boys. The Journal of Social Psychology, 108, 151-159.

Marsh, H. W. (1993). Stability of individual differences in multiwave panel studies: Comparison of simplex models and one-factor models. Journal of Educational Measurement, 30, 157-183.

Marsh, H. W., Balla, J., & Hau, K. T. (1994). An evaluation of incremental fit indices: A clarification of mathematical and empirical properties.

Marsh, H. W., & Hau, K. T. (1994). Assessing goodness of fit: When parsimony is undesirable. (in review)

Marsh, H. W., Hau, K. T., Roche, L., Craven, R., Balla, J., & McInerney, V. (1994). Problems in the application of structural equation modelling: Comment on Randhawa, Beamer, and Lundberg (1993). Journal of Educational Psychology, 86, 457-462.

Randhawa, B. S., & Beamer, J. E. (1994). Structural equation modelling: Reply to Marsh et al. (1994). Journal of Educational Psychology, 86, 463-466.

Randhawa, B. S., Beamer, J. E., & Lundberg, I. (1993). Role of mathematics self-efficacy in the structural model of mathematics achievement. Journal of Educational Psychology, 85, 41-48.

Rest, J. R. (1979). Revised manual for the Defining Issues Test: An objective test of moral judgement development. Unpublished manuscript. Minneapolis, MN: University of Minnesota.

Rest, J. R. (1983). Morality. In J. H. Flavell & E. M. Markman (Eds.), Handbook of child psychology (pp.556-629). NY: John Wiley & Sons.

Rest, J. R. (1986). Moral development: Advances in research and theory. NY: Praeger.

Sachs, J. (1992). Covariance structure analysis of a test of moral orientation and moral judgment. Educational and Psychological Measurement, 52, 825-833.

Snarey, J. R. (1985). Cross-cultural universality of social-moral development: A critical review of Kohlbergian research. Psychological Bulletin, 97, 202-232.

White, C. B., Bushnel, N., & Regnemer, J. L. (1978). Moral development in Bahamian school children: A 3-year examination of Kohlberg's stages of moral development. Developmental Psychology, 14, 58-65.

Table 1

Goodness of Fit and Properness of Alternative Simplex Models

Model[a]	Moral Orientation: Chinese Chi	TLI	RNI	Remarks[b]	Moral Orientation: British Chi	TLI	RNI	Remarks	Moral Judgment: Chinese Chi	TLI	RNI	Remarks	Moral Judgment: British Chi	TLI	RNI	Remarks
Sach's (1992) specification ($q_1 = 0$, $q_4 = 0$)																
M0 (2->3->4->5)	2.23	.96	.99	Imp, TE(-.87)	.46	1.03	1.00	Worry (TE=0.00)	7.83	.93	.99	Worry (TE=0.00)	1.81	.97	.99	Worry TE(0.00)
Present Study ($q_1 = q_2$, $q_3 = q_4$)																
A. Theoretical																
M1 2->3->4->5	2.23	.96	.99	Imp, TE(-.87)	.46	1.03	1.00	Proper	7.83	.93	.99	Worry BE(1.00)	1.81	.97	.99	Proper
B. Minor Reversal																
M2 3->2->4->5	2.23	.96	.99	Imp, PS(-29.10)	.46	1.03	1.00	Imp, BE(1.57)	7.83	.93	.99	Imp, PS(-.44)	1.81	.97	.99	Imp, PS(-.88)
M3 2->4->3->5				Nonconvergence				Nonconvergence	59.42	.39	.90	Imp, PS(-.60)	1.42	.98	1.00	Imp, PS(-.33)
M4 2->3->5->4	2.23	.96	.99	Imp, TE(-.87)	.46	1.03	1.00	Imp, PS(-5.97)	7.83	.93	.99	Worry BE(1.00)	1.81	.97	.99	Imp, PS(-.04)
C. Moderate Reversal																
M5 2->4->5->3				Nonconvergence	2.61	.91	.99	Imp, PS(-35.04)	59.42	.39	.90	Imp, PS(-.17)	1.42	.98	1.00	Proper
M6 3->4->2->5	67.74	-.94	.68	Imp, PS(-21.09)				Nonconvergence	25.96	.74	.96	Imp, PS(-.38)	.01	1.04	1.01	Imp, PS(-1.09)
M7 4->2->3->5				Nonconvergence	2.61	.91	.99	Imp, PS(-1.73)	59.42	.39	.90	Imp, PS(-.60)	1.42	.98	1.00	Imp, BE(1.63)
M8 2->5->3->4				Nonconvergence				Nonconvergence	25.96	.74	.96	Imp, BE(1.44)	.01	1.04	1.01	Imp, PS(-.40)
D. Serious Reversal																
M9 2->5->4->3				Nonconvergence				Nonconvergence	25.96	.74	.96	Imp, BE(1.44)	.01	1.04	1.01	Imp, PS(-.40)
M10 3->4->5->2	67.74	-.94	.68	Imp, PS(-21.09)				Nonconvergence	25.96	.74	.96	Imp, BE(1.44)	.01	1.04	1.01	Imp, PS(-.40)
M11 5->2->3->4	67.74	-.94	.68	Imp, PS(-.48)	3.46	.86	.98	Proper	25.96	.74	.96	Imp, BE(1.44)	.01	1.04	1.01	Imp, PS(-1.09)
M12 3->2->5->4	2.23	.96	.99	Imp, PS(-29.10)	.46	1.03	1.00	Imp, PS(-5.97)	7.83	.93	.99	Imp, PS(-.44)	1.81	.97	.99	Imp, PS(-.88)
M13 4->5->2->3	2.23	.96	.99	Imp, PS(-29.10)	.46	1.03	1.00	Imp, PS(-5.97)	7.83	.93	.99	Imp, PS(-.44)	1.81	.97	.99	Imp, PS(-.88)
M14 4->2->5->3				Nonconvergence	2.61	.91	.99	Imp, PS(-35.04)	59.42	.39	.90	Imp, BE(1.45)	1.42	.98	1.00	Imp, PS(-1.67)
M15 3->5->2->4	50.03	-.42	.76	Imp, BE(4379)				Nonconvergence	59.42	.39	.90	Imp, BE(1.45)	1.42	.98	1.00	Imp, PS(-1.67)
M16 3->5->4->2	50.03	-.42	.76	Imp, BE(3766)				Nonconvergence	59.42	.39	.90	Imp, PS(-.17)	1.42	.98	1.00	Proper
M17 4->5->3->2	2.23	.96	.99	Imp, PS(-.82)	.46	1.03	1.00	Imp, PS(-5.97)	7.83	.93	.99	Worry, BE(1.00)	1.81	.97	.99	Imp, PS(-.04)
M18 4->3->5->2	67.74	-.94	.68	Imp, PS(-7.39)	3.46	.86	.98	Imp, PS(-76.95)	25.96	.74	.96	Imp, BE(1.44)	.01	1.04	1.01	Imp, PS(-.40)
M19 4->3->2->5	67.74	-.94	.68	Imp, PS(-.48)	3.46	.86	.98	Proper	25.96	.74	.96	Imp, BE(1.44)	.01	1.04	1.01	Imp, PS(-1.09)
M20 5->2->4->3	67.74	-.94	.68	Imp, PS(-21.09)	3.46	.86	.98	Imp, PS(-26.58)	25.96	.74	.96	Imp, BE(1.44)	.01	1.04	1.01	Imp, PS(-1.09)
M21 5->3->4->2	50.03	-.42	.76	Imp, BE(1.59)	2.61	.91	.99	Imp, PS(-8.29)	59.42	.39	.90	Imp, PS(-.60)	1.42	.98	1.00	Imp, PS(-.33)
M22 5->3->2->4	50.03	-.42	.76	Imp, BE(2512)	2.61	.91	.99	Imp, PS(-1.73)	59.42	.39	.90	Imp, PS(-.60)	1.42	.98	1.00	Imp, PS(-1.67)
M23 5->4->2->3	2.23	.96	.99	Imp, PS(-29.10)	.46	1.03	1.00	Imp, PS(-1.45)	7.83	.93	.99	Imp, PS(-.44)	1.81	.97	.99	Imp, PS(-.88)
M24 5->4->3->2	2.23	.96	.99	Imp, TE(-.87)	.46	1.03	1.00	Proper	7.83	.93	.99	Worry, BE(1.00)	1.81	.97	.99	Proper

[a] all models have df=1. [b]Proper = solution is proper; Imp = Solution is improper, one source of the improperness is indicated with its out of range value in brackets, PS = y , BE = b , TE = q ; Nonconvergence = solution not converged after 100 iterations.

Table 2
Goodness of Fit and Properness of Some Alternative non-Simplex Models

	Moral Orientation								Moral Judgment							
	Chinese				British				Chinese				British			
Model[a]	Chi	TLI	RNI	Remarks[b]	Chi	TLI	RNI	Remarks	Chi	TLI	RNI	Remarks	Chi	TLI	RNI	Remarks
Theoretical Model																
M0A. ($q_1 = q_2 = q_3 = q_4$)																
2->3->4->5	6.69	.93	.98	Proper	.94	1.03	1.01	Proper	8.03	.97	.99	Imp, PS(-.06)	2.20	1.00	1.00	Imp, PS(-.07)
M0B. ($q_1 = q_2 = q_3 = 1.05\ q_4$)																
2->3->4->5	6.69	.93	.98	Proper	.94	1.03	1.01	Proper	8.03	.97	.99	Proper	2.20	1.00	1.00	Proper
Alternative Non-Simplex Models (all q s equal)																
M1 2->(3,4,5)	67.88	.04	.68	Imp, BE(1.90)	3.78	.95	.98	Imp, BE(1.44)	60.92	.69	.90	Imp, PS(-1.30)	2.17	1.00	1.00	Imp, PS(-1.44)
M2 2->3->(4,5)	67.88	.04	.68	Proper	3.78	.95	.98	Proper	60.92	.69	.90	Imp, PS(-.73)	2.17	1.00	1.00	Imp, PS(-.34)
M3 2->(3,4->5)	4.93	.96	.99	Imp, PS(-2.52)	.47	1.04	1.01	Imp, PS(-1.56)	15.25	.93	.98	Imp, PS(-.36)	6.99	.90	.97	Imp, PS(-.52)
M4 (2,3,4)->5	137.13	-.96	.35	Imp, TE(uniden)	109.10	-1.97	.01	Imp, TE(uniden)	269.85	-.39	.54	Imp, TE(uniden)	71.55	-.43	.52	Imp, TE(uniden)
M5 2->(3->5,4)	53.42	.25	.75	Imp, PS(18.78)	2.64	.98	.99	Imp, PS(-2.00)	60.57	.70	.90	Imp, PS(-.86)	3.27	.97	.99	Imp, PS(-.96)
M6 2->(3->4,5)	72.44	-.02	.66	Imp, BE(1.50)	3.60	.96	.99	Proper	27.73	.87	.96	Imp, PS(-1.05)	1.88	1.00	1.00	Imp, PS(-.72)
M7 (2,3)->4->5	125.57	-.79	.40	Imp, BE(-.06)	101.33	-1.76	.08	Proper	155.65	.20	.73	Imp, Worry, BE(1.00)	21.40	.60	.87	Proper
M8 (2->4,3)->5				Nonconvergence	106.29	-1.89	.04	Proper	187.88	.68	.68	Proper	45.14	.11	.70	Proper
M9 (2->3,4)->5	13.83	.83	.94	Proper	8.23	.83	.94	Imp, PS(-1.07)	122.04	.38	.79	Proper	51.95	-.83	.66	Proper
M10 (2,3->4)->5	124.95	-.79	.40	Proper	100.89	-1.74	.09	Proper	187.88	.04	.68	Imp, PS(-.25)	34.96	.32	.77	Imp, PS(-.17)

[a]For all models, all q s are constrained to be equal (or in M0B, fixed relative to another q), hence df=2.

All models have only 3 paths linking the latent stages. Models separated by commas in brackets are equivalent in the hierarchical order. For example, in M1, stage 2 has three paths to stages 3, 4, and 5 (but there is no path among stages 3, 4 and 5). In M3, stage 2 has two paths to stages 3 and 4; whereas stage 4 has another path to stage 5, but there is no path between stages 3 and 5. In M10, stages 2 and 4 have paths going to stage 5, and stage 3 has another path to stage 4.

[b]Proper = solution is proper; Imp = Solution is improper, one source of the improperness is indicated with its out of range value in brackets; PS = y , BE = b , TE = q ; uniden = unidentified; Nonconvergence = solution not converged after 100 iterations; Worry = worrisome because some of the parameter estimates fall on the boundary of permissible region.

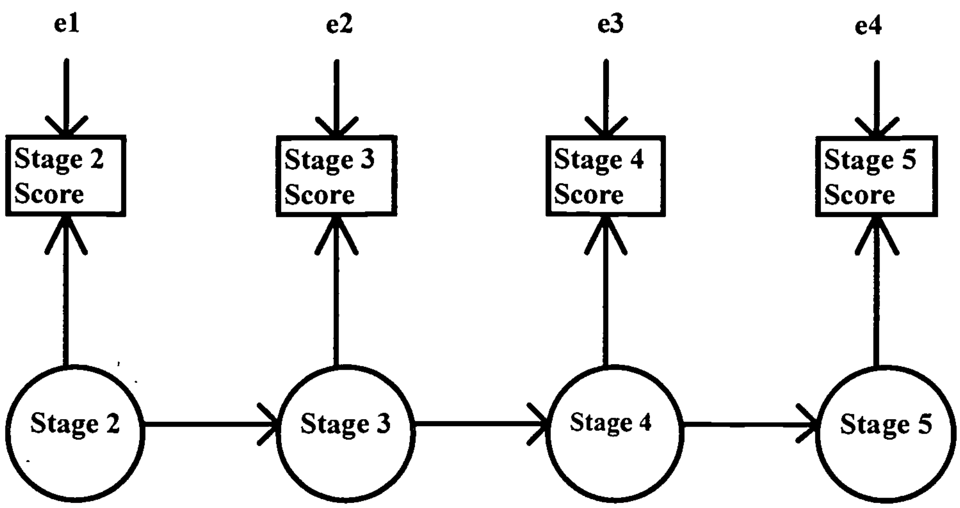

Figure 1

Quasi-Simplex Model of Moral Development

REPRODUCTION RELEASE

(Specific Document)

I. DOCUMENT IDENTIFICATION:

Title: Stage structure of Moral Development: A comparison of alternative models	
Author(s): KIT-TAI HAU	
Corporate Source: The Chinese University of Hong Kong Faculty of Education.	Publication Date: 7 April, 1996

II. REPRODUCTION RELEASE:

In order to disseminate as widely as possible timely and significant materials of interest to the educational community, documents announced in the monthly abstract journal of the ERIC system, *Resources in Education* (RIE), are usually made available to users in microfiche, reproduced paper copy, and electronic/optical media, and sold through the ERIC Document Reproduction Service (EDRS) or other ERIC vendors. Credit is given to the source of each document, and, if reproduction release is granted, one of the following notices is affixed to the document.

If permission is granted to reproduce the identified document, please CHECK ONE of the following options and sign the release below.

[√] Sample sticker to be affixed to document

Sample sticker to be affixed to document [ ]

Check here

Permitting microfiche (4" x 6" film), paper copy, electronic, and optical media reproduction

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY

Hau

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

Level 1

"PERMISSION TO REPRODUCE THIS MATERIAL IN OTHER THAN PAPER COPY HAS BEEN GRANTED BY

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

Level 2

or here

Permitting reproduction in microfiche of other ERIC archival media (e.g. electronic or optical), but not in paper copy.

Sign Here, Please

Documents will be processed as indicated provided reproduction quality permits. If permission to reproduce is granted, but neither box is checked, documents will be processed at Level 1.

"I hereby grant to the Educational Resources Information Center (ERIC) nonexclusive permission to reproduce this document as indicated above. Reproduction from the ERIC microfiche or electronic/optical media by persons other than ERIC employees and its system contractors requires permission from the copyright holder. Exception is made for non-profit reproduction by libraries and other service agencies to satisfy information needs of educators in response to discrete inquiries."

Signature: Hau	Position: Associate Professor
Printed Name: KIT-TAI HAU	Organization: The Chinese University of Hong Kong
Address: Fac of Education The Chinese Univ of Hong Kong, Shatin, NT. HONG KONG	Telephone Number: (852) - 2609 6944
	Date: 2 April, 1996

PS 024236

You can send this form and your document to the ERIC Clearinghouse on Assessment and Evaluation. They will forward your materials to the appropriate ERIC Clearinghouse. ERIC/AERA Acquisitions, ERIC Clearinghouse on Assessment and Evaluation, 210 O'Boyle Hall, The Catholic University of America, Washington, DC 20064, (800) 464-3742